(12) UK Patent Application (19) GB (11) 2 272 442 (13) A

(43) Date of A Publication 18.05.1994

(21) Application No 9323035.7

(22) Date of Filing 09.11.1993

(30) Priority Data
(31) 9223610 (32) 11.11.1992 (33) GB

(71) Applicant(s)
Zeneca Limited

(Incorporated in the United Kingdom)

Imperial Chemical House, 9 Millbank, LONDON, SW1P 3JF, United Kingdom

(72) Inventor(s)
Andrew David Charles

(74) Agent and/or Address for Service
Kevin Bill
Imperial Chemical Industries PLC, ICI Group Patents, Group Patents Services Dept, PO Box 6, Shire Park, Bessemer Road, WELWYN GARDEN CITY, Hertfordshire, AL7 1HD, United Kingdom

(51) INT CL[5]
C12N 9/10 9/12 15/54 , C12Q 1/48 1/68 // (C12N 9/10 C12R 1:125 1:19 1:645 1:865 1:91)

(52) UK CL (Edition M)
C3H HB7E H649 H650 H651 H654 H656 H658 H720
C6Y Y115 Y125 Y330 Y342 Y405 Y406 Y409
U1S S1333 S1334 S1337

(56) Documents Cited
GB 2249099 A
Chemical Abstracts 119(17):176553x & J.Biol.Chem. Vol 268(17) pages 12818-12824,(1993). Biosis Abstract Number:96015754 & Mol.Cell.Biol.,13(5),(1993 pages 2706-2717 Biosis Number:95004297,& J.Biol.Chem. Vol.267 (30), (1992) pages 21368-21374

(58) Field of Search
UK CL (Edition M) C3H HB7E
INT CL[5] C12N 9/10 9/12 15/54
ONLINE DATABASES: WPI,DIALOG(BIOTECH), CAS-ONLINE

(54) Human squalene synthase

(57) Nucleotide sequences, particularly DNA sequences, coding for all or part of human squalene synthase (farnesyl diphosphate farnesyl transferase, EC 2.5.1.21) are described together with vectors containing the DNA sequences, host cells containing the vectors and polypeptides having all or part of the amino acid sequence of human squalene synthase. Processes for preparing the aforementioned vectors, host cells and polypeptides by recombinant DNA technology are also described.

GB 2 272 442 A

Fig.1.

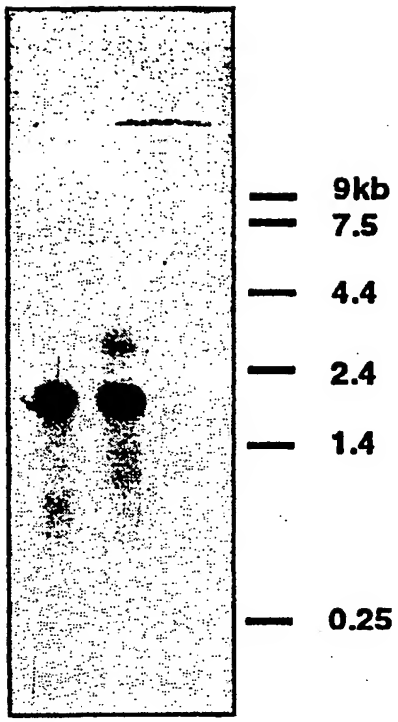


Fig.2

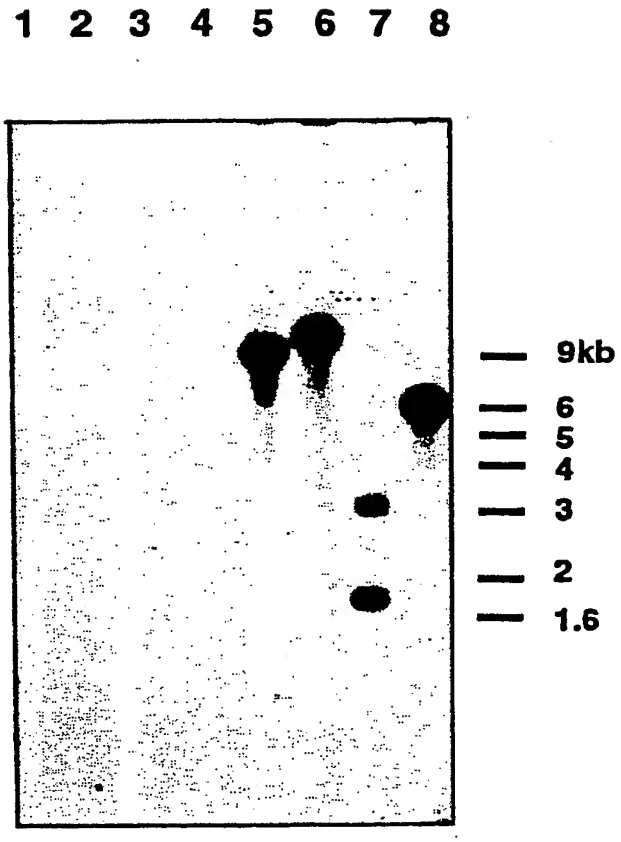


Fig.3.

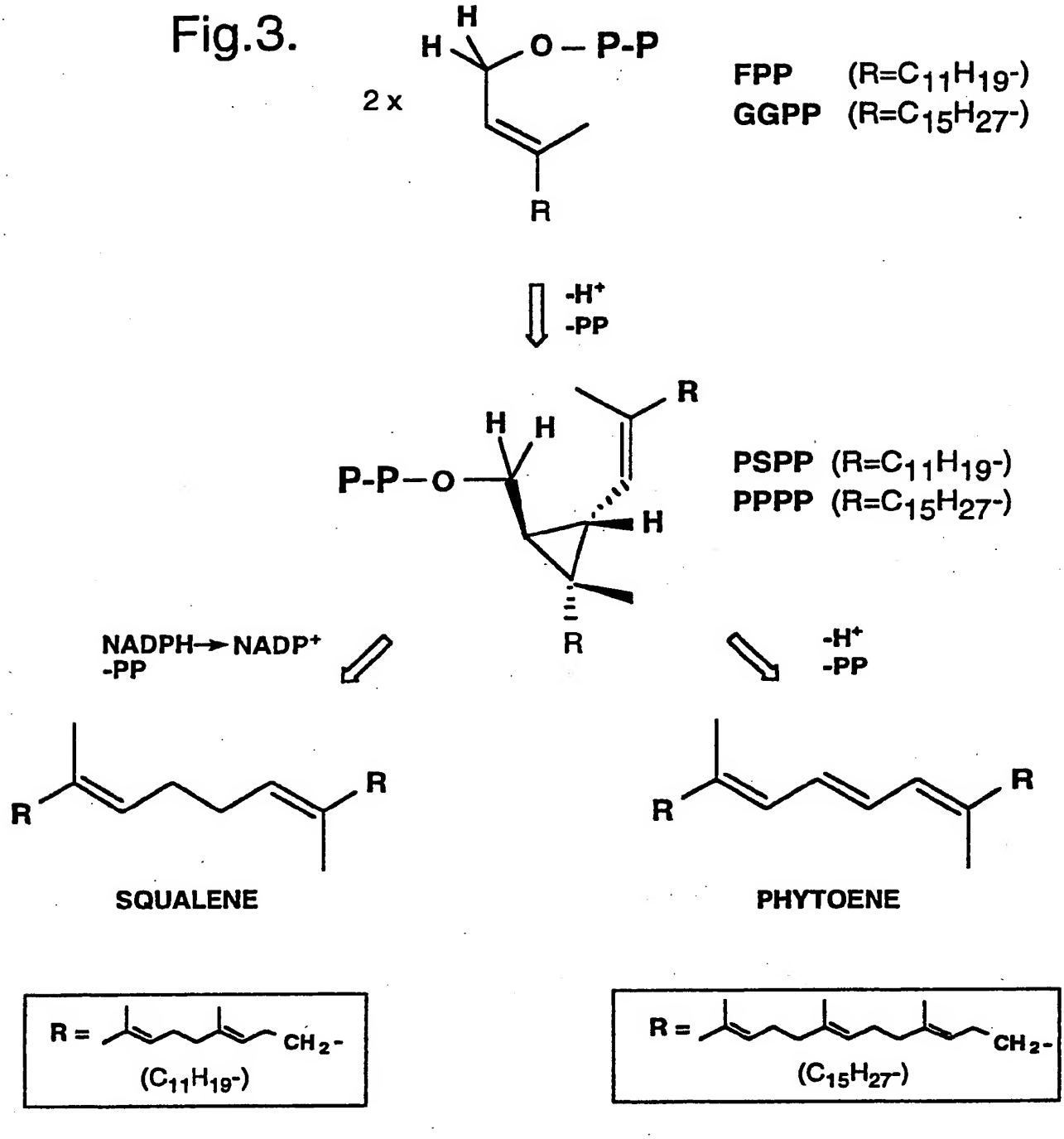


Fig.4.

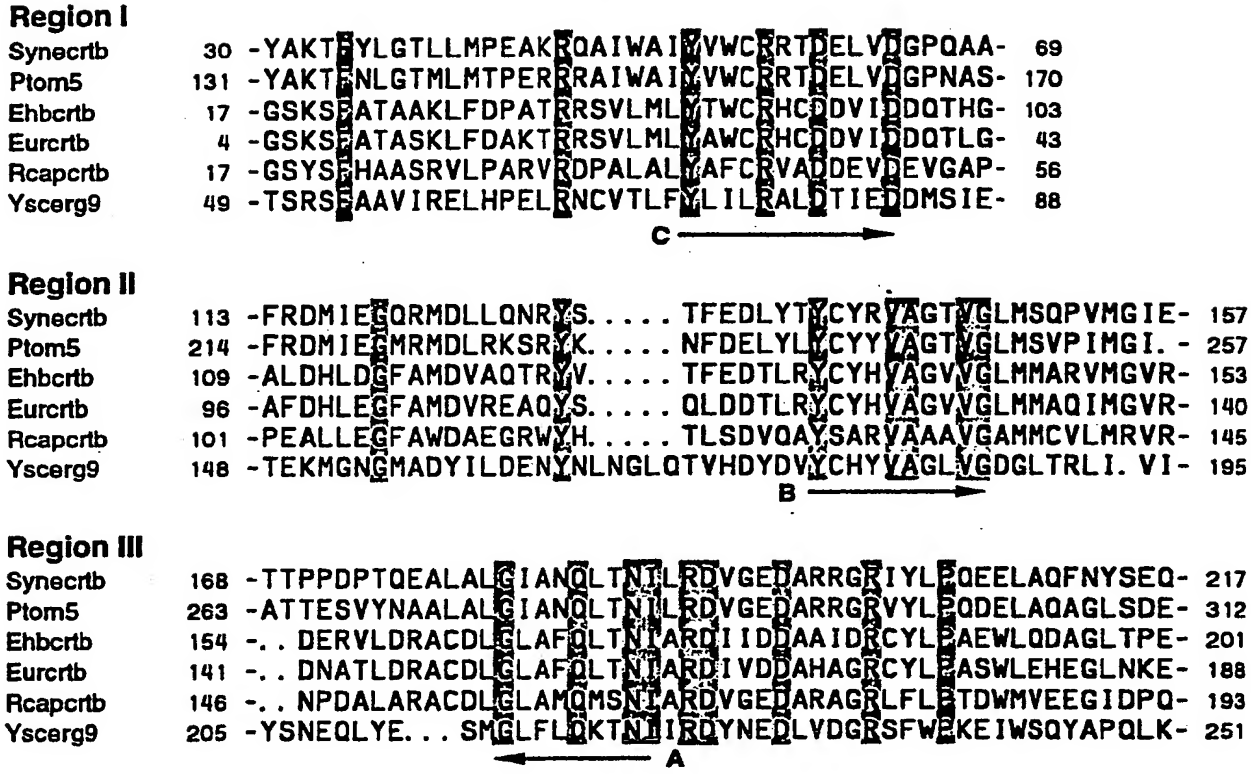


Fig.5.

A

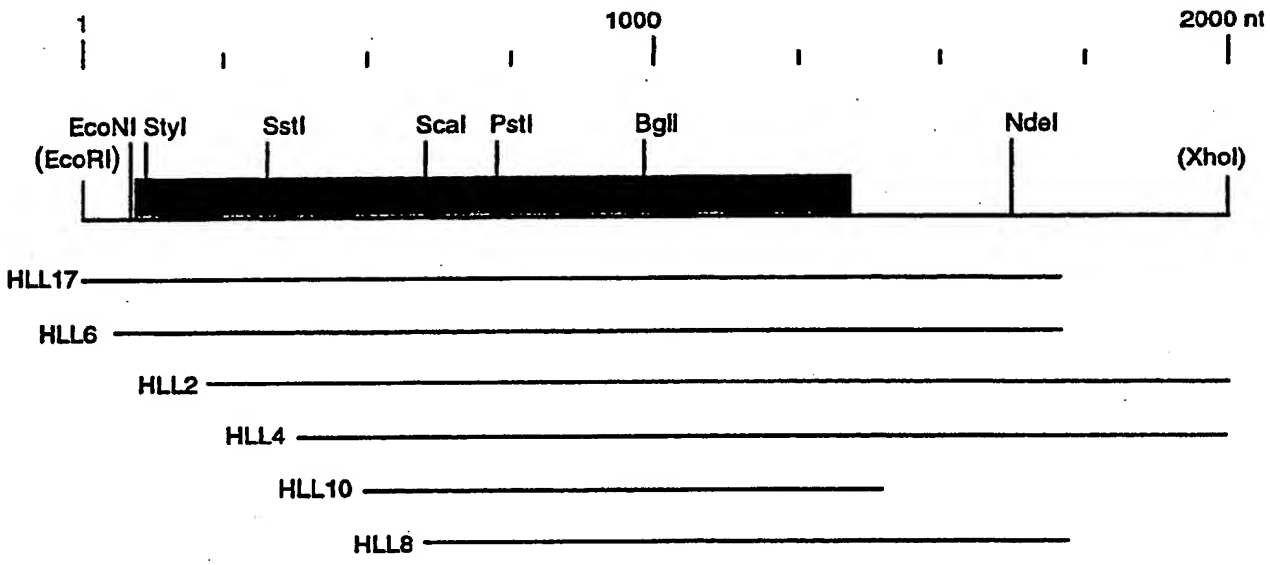
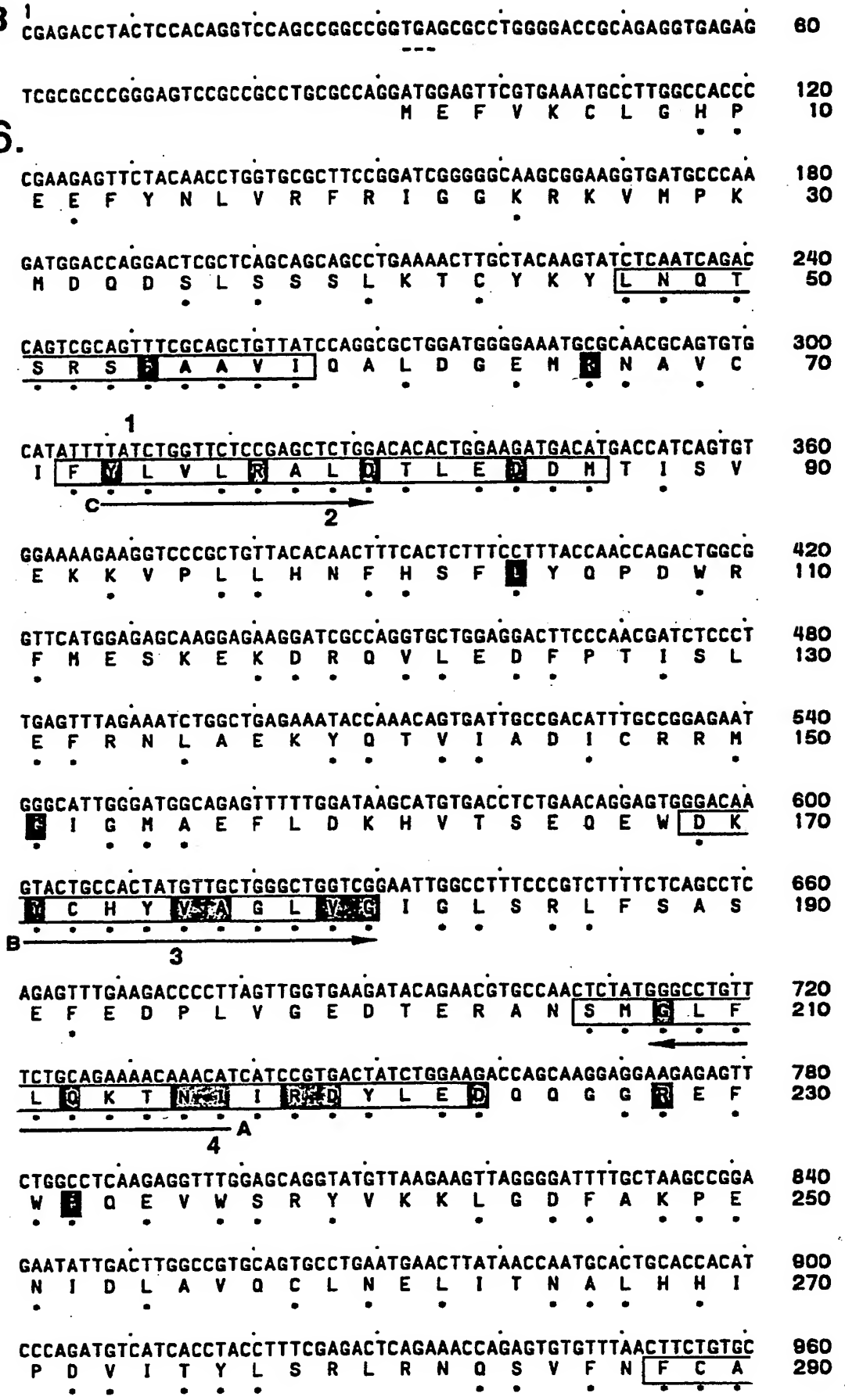


Fig.6.



```
TATTCCACAGGTGATGGCCATTGCCACTTTGGCTGCCTGTTATAATAACCAGCAGGTGTT  1020
 I  P  Q  V  M  A  I  A  T  L  A  A  C  Y  N  N  Q  Q  V  F    310
                5

CAAAGGGGCAGTGAAGATTCGGAAAGGGCAAGCAGTGACCCTGATGATGGATGCCACCAA  1080
 K  G  A  V  K  I  R  K  G  Q  A  V  T  L  M  M  D  A  T  N    330
                6

TATGCCAGCTGTCAAAGCCATCATATATCAGTATATGGAAGAGATTTATCATAGAATCCC  1140
 M  P  A  V  K  A  I  I  Y  Q  Y  M  E  E  I  Y  H  R  I  P    350

CGACTCAGACCCATCTTCTAGCAAAACAAGGCAGATCATCTCCACCATCCGGACGCAGAA  1200
 D  S  D  P  S  S  S  K  T  R  Q  I  I  S  T  I  R  T  Q  N    370

TCTTCCCAACTGTCAGCTGATTTCCCGAAGCCACTACTCCCCCATCTACCTGTCGTTTGT  1260
 L  P  N  C  Q  L  I  S  R  S  H  Y  S  P  I  Y  L  S  F  V    390

CATGCTTTTGGCTGCCCTGAGCTGGCAGTACCTGGCCACTCTCTCCCAGGTAACAGAAGA  1320
 M  L  L  A  A  L  S  W  Q  Y  L  A  T  L  S  Q  V  T  E  D    410

CTATGTTCAGACTGGAGAACACTGATCCCAAATTTGTCCATAGCTGAAGTCCACCATAAA  1380
 Y  V  Q  T  G  E  H  *                                        417

GTGGATTTACTTTTTTTCTTTAAGGATGGATGTTGTGTTCTCTTTATTTTTTTCCTACTA  1440

CTTTAATCCCTAAAAGAACGCTGTGTGGCTGGGACCTTTAGGAAAGTGAAATGCAGGTGA  1500

GAAGAACCTAAACATGAAAGGAAAGGGTGCCTCATCCCAGCAACCTGTCCTTGTGGGTGA  1560

TGATCACTGTGCTGCTTGTGGCTCATGGCAGAGCATTCAGTGCCACGGTTTAGGTGAAGT  1620

CGCTGCATATGTGACTGTCATGAGATCCTACTTAGTATGATCCTGGCTAGAATGATAATT  1680

AAAAGTATTTAATTTGAAGCACCATTTGAATGTTCGTAATAGTAGAAAATGATGTGAATT  1740

TTCTTTCTGTTCGGCTCCTATTTTTCTCATCATTTTGTTTTCTTTAATTGGGTTGAATGG  1800

AGTAGATAGAAATATTTATGGTTTAGGTAACAGTTAGATGTTTCCTAAGAATGCAAACTG  1860

CCTTTTCCACACAAAGGCTGGGAATAAAATTCTGGGTATTCTCGTATTCTCATTTAAAGG  1920

AGTTTAGCTTTCAGAGAGAAACAGCAGGATTGCTTTTGACCTTTTAGAAGATTGGTCTCC  1980

AGTAAAGGTGGACATTTTTGAGATTTTTATAATAAAGAATTTAATTGCTCTGCAAAAAAA  2040

AAAAAAAAAAA  2051
```

Fig.6
(Cont).

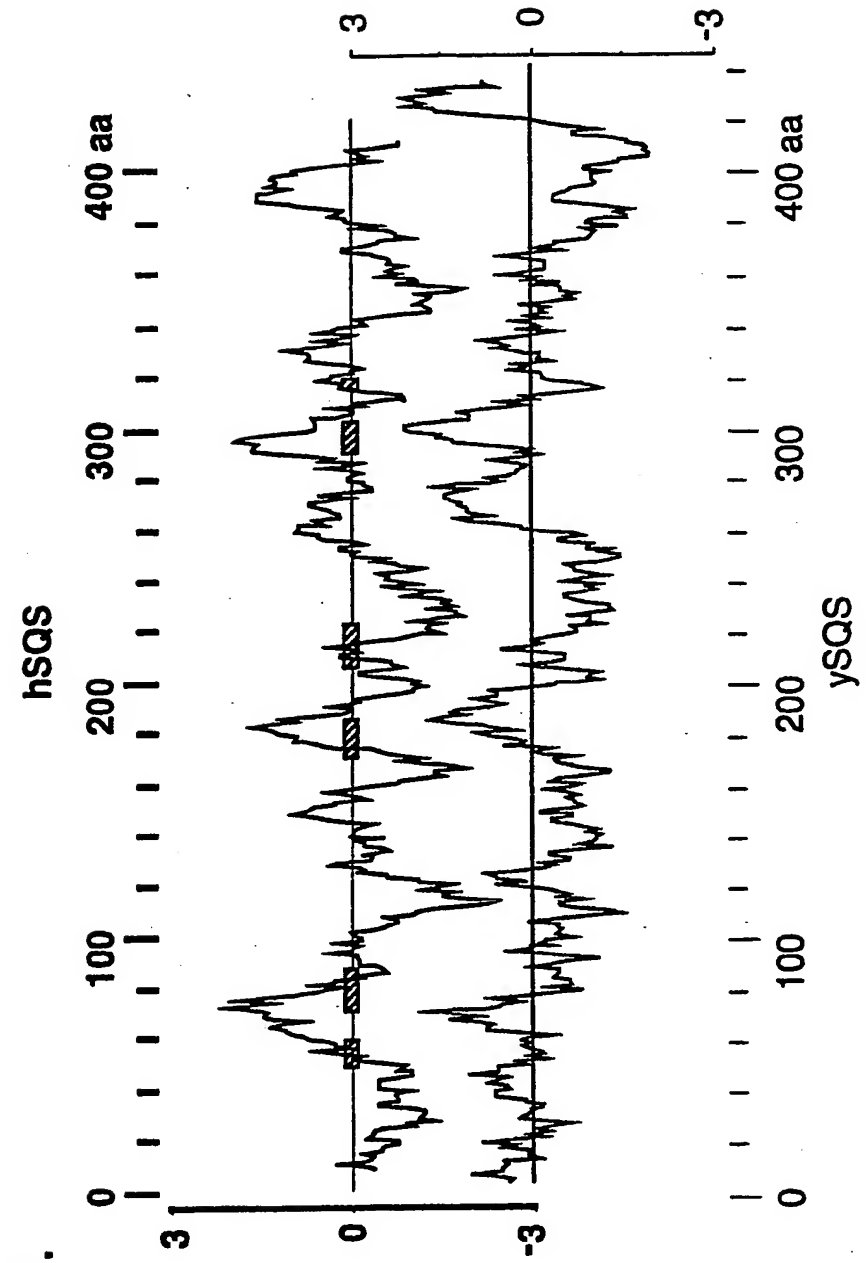


Fig.7.

Fig.8.

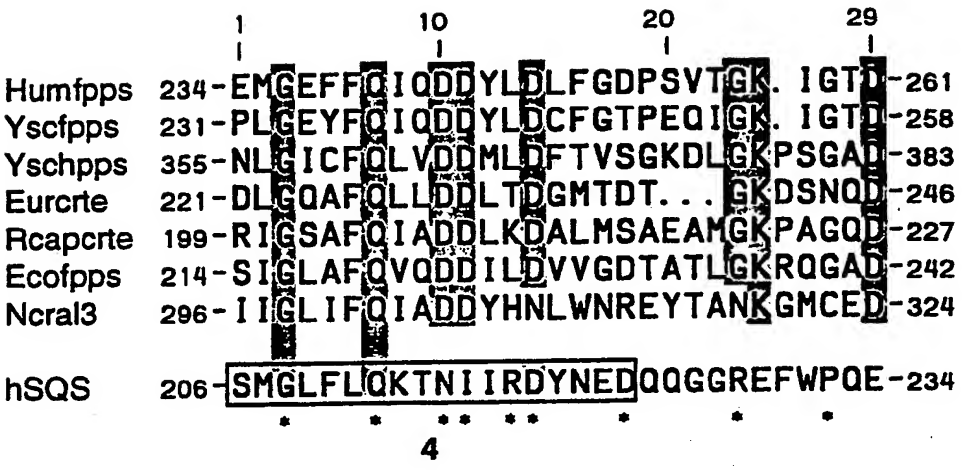
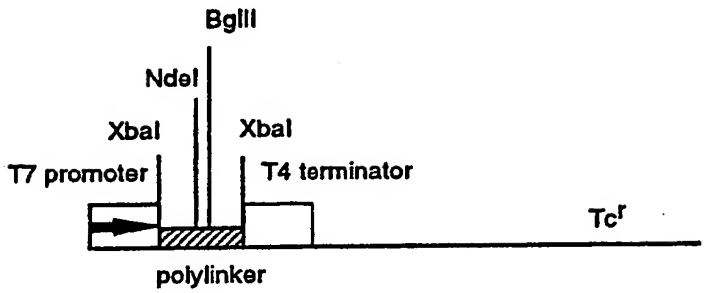
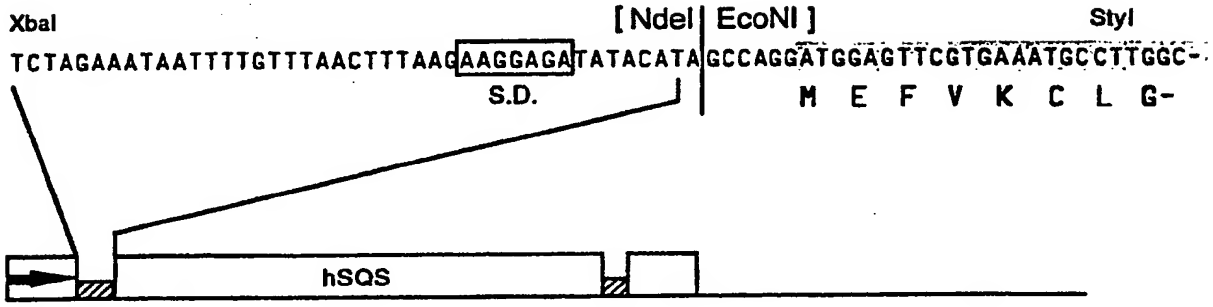


Fig.9.

A. pTB361



B. pHSQS6



C. pHSQS7

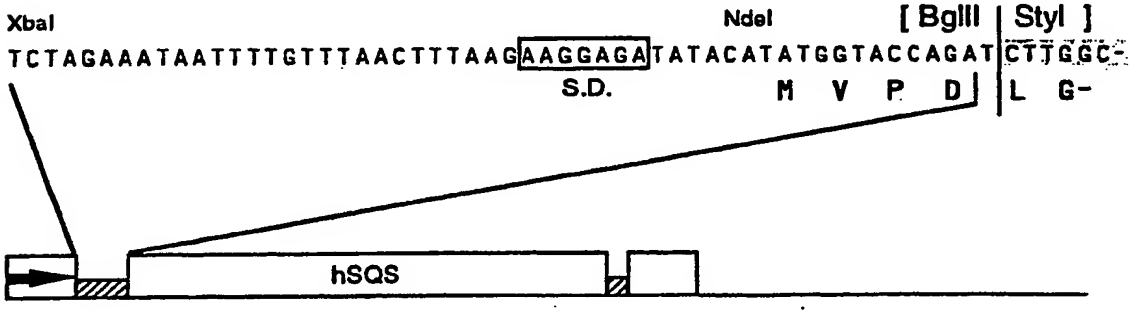
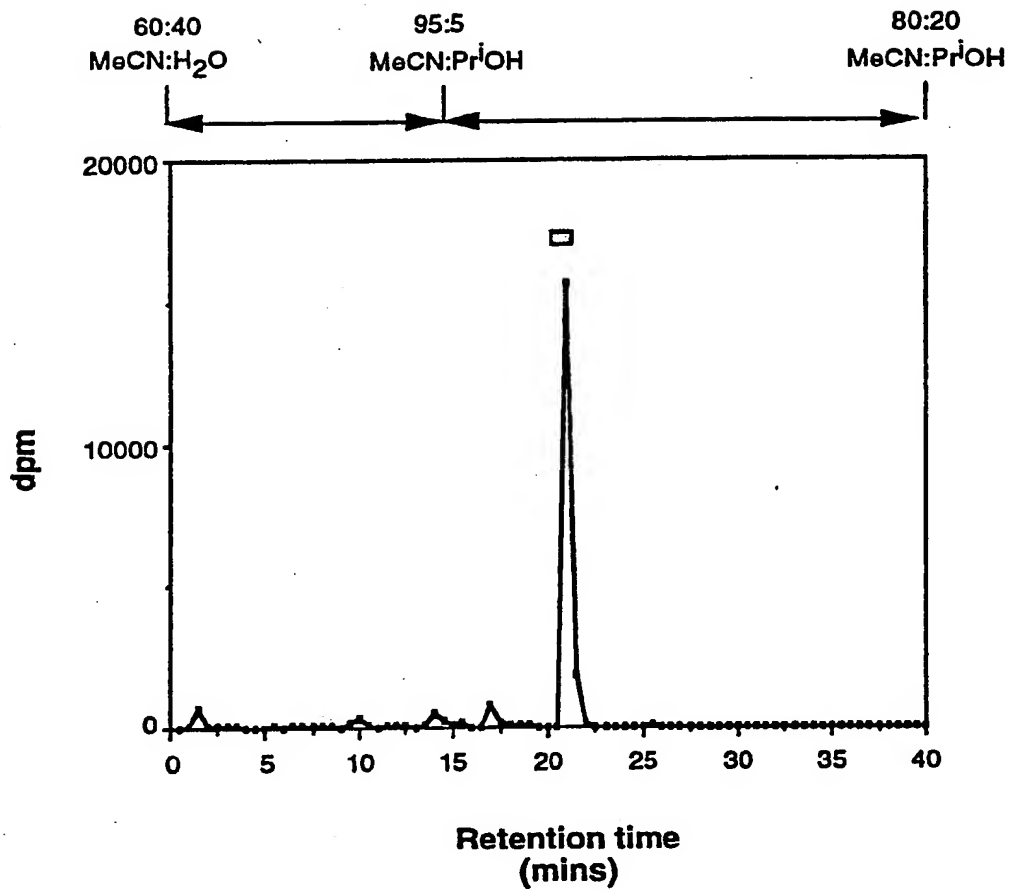


Fig.10.



NUCLEOTIDE SEQUENCES

The present invention relates to human squalene synthase, nucleotide sequences coding therefor and antibodies thereto as well as to processes for the preparation of, and assays based on, human squalene synthase.

Background

Elevated serum cholesterol levels are clearly implicated as a major risk factor in the development of coronary heart disease. One important class of drug that lowers serum cholesterol concentrations is the HMG-CoA (3-hydroxy-3-methyl glutaryl-S-coenzyme A)reductase inhibitors, exemplified by lovastatin (disclosed in U.S. Patent No. 4,231,938). Such compounds inhibit an early step in the cholesterol biosynthetic pathway, namely the conversion of HMGCoA to mevalonate. This serves to lower the intracellular pool of cholesterol, which consequently up-regulates expression of the low density lipoprotein receptor (LDL-R) (Brown M.S. and Goldstein J.L. (1986) Science, 232, 34-47). Elevation of LDL-R levels causes a lowering of the plasma concentration of cholesterol-carrying low density lipoprotein (LDL) particles.

Mevalonate is subsequently converted in a number of steps to the isoprenoid derivative farnesyl pyrophosphate (FPP). FPP is a key intermediate which is required not only for the production of cholesterol, but is also used in the synthesis of other essential cellular products, such as isopentenyl tRNA, ubiquinone, dolichol and the attachment of prenyl groups to membrane localised proteins such as p21-ras (Glomset J.A. et al. (1990) Trends in Biochem., 15, p139-142; Clarke, S. (1992) Annu. Rev. Biochem. 61, 355-386 and references therein). Depletion of some of these other products derived from FPP may have adverse effects on the cell, and thus account for some of the known side-effects associated with HMGCoA reductase inhibitors (McDonald J.S. (1988) Amer. J. Cardiol., 62, 16J-21J).

Squalene synthase (farnesyl diphosphate farnesyl transferase EC 2.5.1.21) is a microsomal enzyme located at the major branch point in the terpene biosynthetic pathway, where the precursors are diverted to one or other specific routes such as long chain isoprenoid or sterol biosynthesis. Squalene synthase catalyses the dimerisation of FPP to squalene in the presence of nicotinamide adenine dinucleotide phosphate (NADPH) (Poulter C.D. and Rilling H.C. (1981) in "Biosynthesis of Isoprenoid Compounds", Vol I, Chapter 8, 413-441, J. Wiley and Sons, and references therin). This enzyme is the first enzyme specific to sterol biosynthesis. Squalene synthase, along with HMG-CoA reductase, is down-regulated by receptor-mediated LDL uptake (Faust J.R et al. (1979) Proc. Natl. Acad. Sci. (USA), 76, 5018-5022), indicating that inhibition of squalene synthase will up-regulate LDL-R expression in a manner similar to that observed with HMGCoA reductase inhibitors. Selective inhibition of this enzyme may therefore lower intracellular cholesterol levels without affecting the synthesis of the other products derived from FPP, and thereby provide an improved therapeutic alternative for the treatment and prevention of hypercholesterolaemia and atherosclerosis.

Squalene synthase has proved difficult to isolate and purify. Most studies to date have utilised crude microsomal preparations (Popjak G. and Agnew W.S. (1979) Mol. Cell. Biochem., 27, 97-116 and references therein; Poulter C.D. and Rilling H.C. (1981) as above, and references therein). A more highly purified preparation of the yeast (Saccharomyces cerevisiae, S. cerevisiae) enzyme has been reported (Sasiak K. and Rilling H.C. (1988) Arch. Biochem. Biophys., 260, 622-627), which contains a 53kD species and a smaller 47kD protein, believed to be derived from the 53kD protein by proteolysis. Very recently, the purification of a soluble squalene synthase activity from rat liver microsomes has been reported, where limited proteolysis was employed to obtain an active enzyme of 32-33kD that could be readily purified in a membrane free fraction (Schechter, I., Klinger, E., Rucker, M. L., Engstrom, R. G., Spiroto, J. A., Islam, M. A., Boettcher, B. R. and Weinstein, D. B. (1992) J. Biol. Chem. 267, 8628-8635).

Recently, the cloning and DNA sequence of the squalene synthase gene from the yeast S. cerevisiae has been reported (Jennings, S. M., Tsay, Y. H., Fisch, T. M. and Robinson, G. W. (1991) Proc. Natl. Acad. Sci. U. S. A. 88, 6038-6042; Fegueur, M., Richard, L., Charles, A. D. and Karst, F. (1991) Curr. Genet. 20, 365-372). In the absence of a sufficiently pure source of squalene synthase to obtain protein sequence information, the gene was cloned by complementation of an erg9 mutant, defective in squalene synthase activity, using genomic yeast DNA fragments. The cloned gene has been used to express the yeast enzyme in both Escherichia coli (E. coli) and S. cerevisiae.

Conventional hybridisation approaches fail to identify human squalene synthase sequences using probes derived from the yeast gene. Other enzymes in the sterol biosynthetic pathway also show limited overall sequence conservation between yeast and humans (30-45%). In the absence of protein sequence derived from purified enzyme, an alternative approach must be found to clone and characterise human squalene synthase, and provide a ready source of the enzyme for kinetic and structural characterisation.

Summary of the invention

The present invention is based, inter alia, on the discovery of nucleotide and amino acid sequences of human squalene synthase.

Thus according to one feature of the present invention there is provided a polypeptide which has all or part of the amino acid sequence and one or more of the biological properties typical of naturally occurring human squalene synthase. The present invention thus encompasses not only allelic variants of human squalene synthase, but also analogues and fragments thereof provided that such analogues and fragments possess the required biological properties.

Preferably the polypeptide has all or part of the amino acid sequence designated SEQ ID No 2 and one or more of the biological

properties typical of human squalene synthase. SEQ ID No 2 depicts the amino acid sequence of human squalene synthase.

In a further embodiment of the present invention we provide a polypeptide having biological properties characteristic of and all or part of the amino acid sequence of human squalene synthase. The polypeptides according to this embodiment thus do not extend to the squalene synthases of species other than human, but do extend to encompass analogues and fragments which possess at least the biological properties of human squalene synthase.

The polypeptides of the present invention may be prepared by chemical synthetic procedures or, more preferably, by procaryotic or eucaryotic host expression. Such expression may be of exogenous DNA sequences for example by bacterial, yeast, higher plant, insect and mammalian cells in culture. The exogenous DNA sequences may be obtained by cloning genomic or complementary DNA or by gene synthesis.

The products of typical yeast (e.g., S. cerevisiae) or procaryote (e.g., E. coli) host cells are free of association with any mammalian proteins. The products of microbial expression in vertebrate (e.g., non-human mammalian and avian) cells are free of association with any human proteins. Depending upon the host employed, polypeptides of the invention may be glycosylated with mammalian or other eucaryotic carbohydrates or may be non-glycosylated. Thus, the polypeptides of the present invention will include human squalene synthase which is in a non-glycosylated form or in a form having other than its natural glycosylation pattern. Polypeptides of the invention may also include an initial methionine amino acid residue (at position -1).

The polypeptides of the present invention may if desired be "labelled" for example by association with a detectable marker, such as a radiolabel, to provide reagents for example for use in an assay for identifying inhibitors of squalene synthase.

According to a further feature of the present invention there is provided DNA sequences coding for all or part of a squalene synthase as hereinbefore defined.

Such sequences may for example include: i) the incorporation of codons "preferred" for expression by selected non-mammalian hosts; ii)

the provision of sites for cleavage by restriction endonuclease enzymes; and iii) the provision of additional initial, terminal or intermediate DNA sequences which facilitate construction of readily expressed vectors. The present invention also provides DNA sequences coding for microbial expression of polypeptide analogues or derivatives of squalene synthase which differ from naturally-occurring forms in terms of the identity or location of one or more amino acid residues (ie. deletion analogs containing less than all of the residues specified for squalene synthase; subsitution analogs, wherein one or more residues specified are replaced by other residues; and addition analogs wherein one or more amino acid residues is added to a terminal or medial portion of the polypeptide) and which share some or all the properties of naturally-occurring forms.

Novel DNA sequences of the invention include sequences useful in securing expression in procaryotic or eucaryotic host cells of polypeptide products having at least a part of the primary structural conformation and one or more of the biological properties of naturally occurring human squalene synthase. DNA sequences of the invention are specifically seen to comprise: (a) the DNA sequence set forth in SEQ ID No 1 or its complementary strand; (b) a DNA sequence which hybridises (under hybridization conditions such as illustrated herein or more stringent conditions) to the DNA sequences in SEQ ID No 1 or to fragments thereof; and (c) a DNA sequence which, but for the degeneracy of the genetic code, would hybridize to the DNA sequence in SEQ ID No 1. Specifically comprehended in part (b) are genomic DNA sequences encoding allelic variant forms of human squalene synthase and/or encoding other mammalian species of squalene synthase. Specifically comprehended by part (c) are manufactured DNA sequences encoding squalene synthase, fragments of squalene synthase and analogs of squalene synthase which DNA sequences may incorporate codons facilitating translation messenger RNA in microbial hosts. Such manufactured sequences may readily be constructed according to the methods of Alton, et al., PCT published application WO 83/04053.

The DNA sequences of the present invention include those useful in securing expression in procaryotic or eucaryotic host cells and the polypeptides of the present invention may be in either glycosylated or non-glycosylated form depending upon the host cell selected. Where the

derivative of the present invention is obtained in non-glycosylated form, for example following expression in procaryotic host cells, the derivative may, if desired, be glycosylated chemically for example with mammalian or other eucaryotic carbohydrates.

According to a further feature of the present invention there is provided a recombinant vector containing a DNA sequence as hereinbefore defined. The recombinant vector may for example be a biologically functional plasmid or viral DNA vector.

According to a further feature of the present invention there is provided a process for the preparation of a recombinant vector as hereinbefore defined which comprises inserting a DNA sequence as hereinbefore defined into a vector.

According to a further feature of the present invention there is provided a procaryotic or eucaryotic host cell stably transformed or transfected with a recombinant vector as hereinbefore defined.

According to a further feature of the present invention there is provided a process for the preparation of a procaryotic or eucaryotic host cell as hereinbefore defined which comprises transforming or transfecting a procaryotic or eucaryotic cell with a recombinant vector as hereinbefore defined whereby to yield a stably transformed or transfected procaryotic or eucaryotic host.

According to a further feature of the present invention there is provided a process for the preparation of a squalene synthase of the present invention which comprises culturing a procaryotic or eucaryotic host cell of the invention whereby to obtain said derivative. The process will advantageously also include the step of isolating the said derivative produced by expression of the DNA sequence of the invention in the recombinant vector of the invention.

The host cells for use in processes of the present invention are preferably procaryotic such as E. coli or Bacillus subtilis, but may be yeast cells such as S. cerevisiae or Schizosaccharomyces pombe (S. pombe), insect cells such as Spodoptera frugiperda or mammalian cells such as CHO cells (Chinese hamster ovary cells) or MEL (murine erythroleukaemia cells).

The recombinant squalene synthase prepared according to the present invention may be used in an assay to detect inhibitors of human

squalene synthase. Thus according to the present invention there is provided the use of a recombinant squalene synthase (as herein defined) to detect inhibitors of human squalene synthase. As mentioned above the squalene synthase will preferably comprise all or part of the amino acid sequence designated in SEQ ID. No. 2. In a particular embodiment the method comprises the preparation of cell extracts containing particulate or soluble cell material from a procaryotic or eucaryotic host producing a squalene synthase by expression of all or part of a human squalene synthase gene of the present invention as hereinbefore defined, and the use of such extracts to measure the inhibitory effects of compounds on the activity of the squalene synthase. This inhibitory effect may be measured by techniques familiar to anyone with ordinary skill in the art, for example, the methods described in published patent applications EP 475,706 and WP 92/15579.

The host cell capable of expressing squalene synthase may if desired be a host cell as hereinbefore described and the gene present therein may be either the endogenous squalene synthase gene, in the case of a mammalian host cell, or a recombinant vector of the present invention as hereinbefore described.

Brief Description of the Drawings

Figure 1. Northern blot analysis of polyA$^+$ RNA from ERG9 (Lane 1) and erg9 (Lane 2) yeast strains, and HepG2 cells (Lane 3)

Figure 2. Southern blot analysis of human (Lanes 1-4) and yeast (Lanes 5-8) genomic DNA, digested with HindIII (lanes 1,5), EcoRI (Lanes 2,6), EcoRV (Lanes 3,7) and BamHI (Lanes 4,8).

Figure 3. Reactions catalysed by squalene synthase (R=$C_{11}H_{19}$-) and phytoene synthase (R=$C_{15}H_{27}$-).

Figure 4 Regions of homology between yeast squalene synthase (Yscerg9) and phytoene synthases from Synechococcus PCC7942 crtB gene (Synecrtb) (Chamovitz, D., Misawa, N.,

Sandmann, G. and Hirschberg, J. (1992) FEBS Letters 296, 305-310), tomato (Ptom5) (Ray, J., Bird, C. R., Maunders, M. J., Grierson, D. and Schich, W. (1987) Nucleic Acids Res. 15, 10587), Erwinia herbicola crtB gene (Ehbcrtb) (Armstrong, G. A., Alberti, M. and Hearst, J. E. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 9975-9979), Erwinia uredovora crtB gene (Eurctrb) (Hisawa, N., Nakagawa, M., Kobayashi, K., Yamano, S., Izawa, Y., Nakamura, K. and Harashima, K. (1990) J. Bacteriol. 172, 6704-6712) and Rhodobacter capsulatus crtB gene (Rcapcrtb) (Armstrong, G. A., Alberti, M. and Hearst, J. E. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 9975-9979). Identical residues are highlighted by shaded boxes.

Figure 5 Structure the cDNA for human squalene synthase. Restriction map of human squalene synthase cDNA , showing the position of the pHLL cDNA clones isolated and characterised. The solid bar indicates the position of the 417 aa ORF. The scale above the map is in bp.

Figure 6 Sequence of the cDNA for human squalene synthases. Sequence of human squalene synthase cDNA, showing the deduced amino acid sequence (single letter code) in the ORF (GenBank/EMB2 accession number X69141). The in-frame stop codon at nt 32 is shown by the dotted underline. The two potential polyadenylation signals (AATAAA) at nt 1883 and nt 2011 are single underlined. The alternative polyadenylation signal ATTAAA at nt 1678 is double underlined. Residues sharing identity between human and yeast squalene synthase are marked with asterisks.

Regions of conservation between the amino acid sequences of the yeast and human enzymes (homology > 85%) are boxed. Regions 2, 3 and 4 correspond to the conserved sequences I, II and III respectively identified in Figure 4. Amino acids which are conserved in phtyoene synthase are

highlighted in filled boxes. The arrowed lines underneath boxes 2, 3 and 4 correspond to the position of the PCR primers C, B and A respectively (see Example 2).

Figure 7. Hydropathy plot of human (top trace) and yeast (lower trace) squalene synthase generated using the algorithm of Kyte and Doolittle (Kyte, J. and Doolittle, R. F., (1982) J. Mol. Biol., 157, 105-132) using a window size of 15, as implemented in the University of Wisconsin GCG software package (Devereux, J. et al., (1984) Nucleic Acids Res, 12, 387-395.

Figure 8 Comparison of the sequence of human squalene synthase in region 4 with polyisoprenoid diphosphate synthases. Region 4 of human squalene synthase is aligned with the aspartate-rich domain II (Ashby, M. N. and Edwards, P. A. (1990) J. Biol. Chem. 265, 13157-13164) from human FPP synthase (Humfpps) (Wilkins, D. J., Kutsunai, S. Y. and Edwards, P. A. (1990) J. Biol. Chem. 265, 4607-4614), S. cerevisiae FPP synthase (Yscfpps) (Anderson, M. S., Yarger, J. G., Burck, C. L. and Poulter, C. D. (1989) J. Biol. Chem. 264, 19176-19184), S. cerevisiae hexaprenyldiphosphate synthase (Yschpps) (Ashby, M. N. and Edwards, P. A. (1990) as above), Erwinia uredovora crtE protein GGPP synthase (Eurcrte) (Misawa, N., Nakagawa, M., Kobayashi, K., Yamano, S., Izawa, Y., Nakamura, K. and Harashima, K. (1990) J. Bacteriol. 172, 6704-6712), Rhodobacter capsulatus crtE protein GGPP synthase (Rcapcrte) (Armstrong, G. A., Alberti, M. and Hearst, J. E. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 9975-9979), E. coli FPP synthase (Ecofpps) (Fujisaki, S., Hara, H., Nishimura, Y., Horiuchi, K. and Nishino, T. (1990) J. Biochem. (Tokyo) 108, 995-1000), and Neurospora albino 3 gene product GGPP synthase (Ncral3) (Carattoli, A., Romano, N., Ballario, P., Morelli, G. and Macino, G. (1991) J. Biol. Chem. 266, 5854-5859). Amino acid

identities are highlighted in shaded boxes. The amino acids in human squalene synthase that are invariant in phytoene synthase are marked with asterisks.

Figure 9. Construction of expression vectors pHSQS6 and pHSQS7 from the vector pTB361.
(A) Structure of pTB361 showing the T7 promoter sequences (arrowed box), the polylinker (hatched box) and the T4 terminator sequences (shaded box). (B) Construction of pHSQS6. The 1.65kb EcoNI-XhoI fragment from pHLL6 was subcloned into pTB361 between the NdeI, and XhoI sites. The resulting construct encodes the complete NH2-terminus of human squalene synthase (shaded sequence). The position of the Shine-Dalgarno sequence in the vector sequences is shown by the open box located 14nt 5′ to the ATG codon. (C) Construction of pHSQS7. The 1.65kb StyI-XhoI fragment from pHLL6 was cloned between the BglII and XhoI sites in pTB361, such that the human squalene synthase sequences (shaded) were fused in frame with a short vector encoded leader sequence. The Shine Dalgarno sequence in pHSQS7 is maintained at its optimal distance of 7 nt from the ATG codon.

Figure 10. HPLC analysis of Fraction 2. Fraction 2 was recovered from a C18 column, concentrated by drying, loaded onto a Spherisorb S30DS1 HPLC column and eluted as described in Example 3. Samples were collected every 30 seconds and radioactivity measured in a scintillation counter. The shaded box marks the retention time for pure squalene. The major radiolabelled component comigrates with squalene, and coincides with a beak in the UV absorbance of the eluate at 210 nm The solvent gradients used are indicated above the Panel.

Detailed Description

A DNA sequence encoding all or part of human squalene synthase has been isolated and characterised.

The present invention demonstrates that the sequence differences between yeast and human squalene synthase are such that the use of fragments of the yeast squalene synthase gene as probes fails to detect homologous human RNA or DNA sequences, using Northern or Southern hybridisation techniques. Such probes can contain either all or part of the complete yeast gene, including fragments embracing the putative isoprenoid binding site (Jennings, S. M. et al. (1991) as above; Ashby, M. N. and Edwards, P. A. (1990) J. Biol. Chem. 265, 13157-13164). This failure derives in part from differences in the amino acid sequences between the proteins, and also due to the difference in usage of preferred codons in the coding sequence within the messanger RNA.

The present invention describes a method for identifying particular amino acid sequences within yeast squalene synthase which would be predicted to be conserved in the human enzyme, even when the surrounding sequences exhibit considerable divergence. These sequences can be used to design oligonucleotide primers in which the codon usage is optimised for human RNAs. These primers can in turn be used to amplify desired DNA sequences from human RNA using techniques familiar to anyone skilled in the art. The method relies on the observation that the reaction catalysed by squalene synthase is mechanistically similar to that undertaken by phytoene synthases in plants and some bacteria. Here we describe how three specific sequences that are conserved between phytoene synthases of different species can also be identified in yeast squalene synthase. We show here that these sequences can be used to identify and clone the cDNA for human squalene synthase.

We further describe hereinafter the introduction of sequences specifying all or part of human squalene synthase carried on an expression vector into a procaryotic host (*E. coli*), and the production of extracts from cells expressing the recombinant human squalene synthase containing soluble forms of the enzyme. These extracts are shown to contain a soluble squalene synthase activity. Such extracts can be used to screen for inhibitory compounds which are selective for human squalene

synthase. This method therefore avoids the necessity for purification of the human enzyme or its reconstitution into membranes for assay purposes.

All or part of the human sequence may be used, if desired, to generate antibodies directed against human squalene synthase.

The present invention is illustrated by the following Examples, in which the materials employed and the techniques used are set out below:-

Sequenase(TM) sequencing kit

A kit comprising 1) a concentrated buffer of sodium chloride, Tris and magnesium chloride for sequencing reactions; 2) a labelling mix containing the deoxynucleotide triphosphate solutions dGTP, dCTP, dTTP; 3) termination mixes containing a concentrated solution of sodium chloride, all four deoxynucleotide triphosphate solutions and one each of the dideoxy nucleotide triphosphate solutions ddATP, ddGTP, ddCTP, ddTTP; 4) sequence extending mixes containing a concentrated solution of sodium chloride and all four deoxynucleotide triphosphates; 5) enzyme dilution buffer containing Tris, dithiothreitol and bovine serum albumin; 6) reaction stop solution containing formamide, and EDTA and 7) a chemically modified T7 DNA polymerase, based on the procedure of Tabor and Richardson (1987) Proc. Nat. Acad. Sci (USA), 84, p4767-4771.

Bacteriophage lambda human liver cDNA library

A population of recombinant bacteriophage λ carrying complementary DNA inserts derived from a human liver polyA$^+$ RNA population by reverse transcription using an oligo(dT) primer. Many such libraries are available from commercial sources. In particular, we have used a modified λ vector system (λZAP) available from Stratagene (La Jolla, California, USA), in which the cDNA inserts may be recovered from the recombinant parental phage in the form of a plasmid by excision with a helper bacteriophage, as described in Short J. M. et al, (1988) Nucleic Acids Res. 16, 7583-7600.

BluescriptII(TM) KS and SK vectors

A recombinant cloning vector system similar to that described by Yanisch-Perron C. et al. (1985) Gene, 33, p109-119, comprising a

ColEI-based replicon bearing a polylinker DNA fragment containing multiple unique restriction sites, flanked by bacteriophage T3 and T7 promoter sequences; a filamentous phage origin of replication and an ampicillin drug resistance marker gene.

Tergitol NP40(TM)

A polyglycol ether (nonionic) surfactant.

Hybond-N(TM)

A supported nylon-66 membrane with a pore size of 0.45 microns, used for the immobilisation of nucleic acids by either u.v. cross linking or oven baking and supplied by Amersham International plc, Amersham, Bucks, UK.

DNA molecular weight standards

A mixture of double-stranded DNA fragments supplied by Gibco-BRL, Paisley, Renfrewshire, Scotland. The mixture comprises the following sizes of fragment (kb); 12.216, 11.198, 10.18, 9.162, 8.144, 7.126, 6.108, 5.09, 4.072, 3.054, 2.036, 1.636, 1.108, 0.517, 0.506, 0.396, 0.344, 0.298, 0.222. 0.201, 0.154, 0.134, 0.075.

RNA molecular weight standards

A mixture of synthetic single-stranded poly-A tailed RNA fragments supplied by Gibco-BRL, Paisley, Renfrewshire, Scotland. The mixture comprises the following sizes of fragment (kb); 9.49, 7.46, 4.40, 2.37, 1.35, 0.24.

SSC

0.15M NaCl + 0.015M sodium citrate pH7.0

Denhardt's reagent

A solution containing 0.02% bovine serum albumin, 0.02% Ficol 400,000 (a nonionic synthetic polymer of sucrose, dialysed and lyophilised and having an approximate molecular weight of 400,000) and 0.02% polyvinyl pyrrolidone

Invitrogen FastTrack(TM) mRNA isolation kit

A kit containing an SDS-based lysis buffer and reagents for the isolation of polyA$^+$ mRNA by oligo-dT cellulose chromatography, essentially as described in "Molecular Cloning - a laboratory manual", Second Edition, Sambrook, Fritsch and Maniatis (Cold Spring Harbour Laboratory, Cold Spring Harbour, New York, USA, 1989), and supplied by Invitrogen Corp, San Diego, California, USA.

First-strand cDNA synthesis kit

A kit containing murine reverse transcriptase, RNase inhibitors, oligo$(dT)_{12-18}$ and nucleotide triphosphates in an aqueous buffer, esentially as described in Sambrook et al, (1989) as above, and supplied by Pharmacia, Milton Keynes, Bucks, UK.

Taq polymerase

A thermostable DNA polymerase isolated from the thermophilic bacterium Thermus aquaticus, and available from commercial sources.

Strains and culture conditions

Bacterial strains were grown in complete or selective medium as described by Yanisch-Perron et al (1985) as above; Sambrook et al. (1989) as above; and Miller J.H. (1972) "Experiments in Molecular Genetics", Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA.

The E. coli T7 RNA polymerase expression strain BL21(DE3) has been described (Studier W. F. and Moffatt B. A. (1986) J. Mol. Biol., 189, 113-130). The T7 expression vector pTB361 contains a bacteriophage T7 promoter and Shine-Dalgarno sequence upstream of a polylinker seqence containing an ATG translational initiation codon, and a bacteriophage T4 transcription termination sequence (see Figure 9 for details).

Plasmid DNA was made from a strain carrying the T7 promoter vector pRK172 (Studier see, for example, J. Mol. Biol., 189, (1986), 113-130). This was cut with XhoII and the promoter region was isolated as a 103 bp fragment. The 5′ overhang ends of the purified fragment were then filled in by Klenow polymerase reaction. The fragment was then cut with NdeI and cloned into pLB015 (see published UK patent application GB2253852)) carrying the trp promoter after

being cut with SspI and NdeI. This effects replacement of the trp promoter with the T7 promoter. A new synthetic multicloning region containing sites for NdeI, KpnI, BglII, NsiI, XhoI, ScaI, XbaI and SalI was then cloned between the T7 promoter and the T4 terminator to generate pTB361.

Yeast (Saccharomyces cerevisiae) strain DB746 (MATα, leu2-3,2-113, his3Δ1, trp1-289, ura3-52, can1) is deposited with the American Type Culture Collection, 12301 Parklawn Drive, Rockville, Maryland, USA under the deposit number ATCC 44773. Recombinant strain FK5188 (MATa, erg9, aux32, ura3) is a derivative of wild type yeast FL100, deposited with the ATCC as deposit number 28383, which is auxotrophic for ergosterol and uracil (Karst F. and Lacroute F. (1977) Mol. Gen. Genet., 154, 259-277) and defective in squalene synthetase and OMP decarboxylase activities (Fegueur, M. et al. (1991) Curr. Genet. 20, 365-372). The aux32 mutation is required for viability of strains bearing the erg9 mutation (Karst F. and Lacroute F. (1977) as above). Strain DB746 was cultured in YPAD medium as described in "Guide to Yeast Genetics and Molecular Biology", C Guthrie, G. R. Fink (eds), (1991), Academic Press, New York, USA. FK5188 was grown in YPAD medium supplemented with 2μg/ml ergosterol in 1:1 Tergitol:ethanol.

Human hepatocellular carcinoma HepG2 cells (ATCC deposit number HB8065) were cultured in RPMI-Dutch Mod. (Northumbria Biologicals Ltd), 1mM sodium pyruvate, 2mM glutamine, 1 x non-essential amino acids, and 10% foetal calf serum. For experiments involving lovastatin, the foetal calf serum was replaced with 10% human lipoprotein-depleted serum (LPDS). Cells were cultured in medium + LPDS for 24h prior to addition of 1 mM lovastatin (sodium salt), and incubated for a further 48h before harvesting. Control cells were incubated in medium + LPDS without added lovastatin.

Preparation of mRNA

Poly A^+ mRNA was prepared directly from ~ 1 x 10^8 HepG2 cells using a FastTrack mRNA isolation kit (Invitrogen). Rat tissue total mRNA was prepared by polytron homogenisation in 4M guanidine isothiocyanate, 2.5mM sodium citrate, 0.5% Sarcosyl, 100mM

β-mercaptoethanol, followed by centrifugation through 5.7M CsCl, 25mM sodium acetate at 135,000g(max). Rat polyA$^+$ mRNA was obtained from total RNA using the FastTrack kit.

Transformations

E. coli transformations were generally carried out by electroporation. 400ml cultures of strains DH5α or BL21(DE3) were grown in L-broth to an OD600 0.5 and harvested at 2,000g. The cells were washed twice in ice-cold deionised water, resuspended in 1ml 10% glycerol and stored in aliquots at -70°C. Ligation mixes were desalted using Millipore 'V' series membranes (0.025mm pore size). 40µl of cells were incubated with 1µl of ligation mix or plasmid DNA on ice for 10 minutes in 0.2cm electroporation cuvettes, and then pulsed using a Gene Pulser apparatus (Biorad) at 6.5kV cm-1, 25µF, 250Ω. Transformants were selected on L-agar supplemented with tetracycline at 10µg/ml or ampicillin at 100µg/ml

Northern blots

2 µg polyA$^+$ or 20 µg total RNA samples were resolved by electrophoresis on 1% denaturing formaldehyde agarose gels in MOPS buffer (Sambrook et al. (1989) as above) and transferred onto Hybond-N (Amersham International, Amersham, UK). Probes were labelled with ^{32}P by random hexamer priming, and hybridisations were carried out in 0.28 M sodium phosphate (pH 7.2), 5 x Denhardt's solution, 10% dextran sulphate, 0.1% SDS at 65°C. Membranes were washed to a final stringency of 0.2 x SSC, 0.1% SDS at 65°C. To control for loading variations, blots were stripped after autoradiography by boiling in 0.1% SDS, and then rehybridised using a probe containing 1.2 kb of a rat glyceraldehyde-3-phosphate dehydrogenase cDNA (G3PD) (Fort, P., Marty, L., Piechaczyk, M., El Sabrouty, S., Dani, C., Jeanteur, P. and Blanchard, J. M. (1985) Nucleic Acids Res. 13, 1431-1442). Analysis of polyA$^+$ RNAs from human tissues was carried out using a panel of commercially available pre-blotted RNAs (Clontech Laboratories, Palo Alto, California, USA). RNA molecular weight standards were obtained from Gibco-BRL as hereinbefore described.

Southern blots

Yeast genomic and total RNA were isolated using published procedures (in "Yeast: a practical approach", (1988) Campbell I. and Duffus J.H. (eds) IRL Press, Oxford). Human genomic DNA was isolated from 1 x 10^7 HepG2 cells by lysing cells in a buffer containing 20mM Tris pH8.0, 10mM NaCl, 1mM EDTA, 100µg/ml proteinase K and 0.5% SDS, extracting with 1:1 phenol:chloroform and precipitation. DNA probes were labelled with ^{32}P by random priming as described. DNA molecular weight standards were obtained from Gibco-BRL (as hereinbefore described). For Southern blots, 5µg of DNA was digested and fragments separated on an 0.6% agarose gel, then transferred to Hybond-N(TM) membrane (Amersham). Probe hybridisation was carried out using the method of Church and Gilbert (Church G.M. and Gilbert W. (1984) Proc. Natl Acad. Sci (USA), 81, p1991-1995) in 0.5mM sodium phosphate, 1mM EDTA, 7% SDS (pH7.2), at 42°C and the membrane was washed in 50mM sodium phosphate, 1% SDS (pH7.2) at 60°C.

Sequencing methods.

DNA sequence analysis was performed by the dideoxy-chain termination of Sanger F. et al. (1977) Proc. Natl. Acad. Sci. (USA), 74, 5463-5467 using Sequenase(TM) and the associated DNA sequencing kit purchased from U.S. Biochemicals (Cleveland, Ohio, USA).

Nomenclature

Nucleic acid and amino acid nomenclature used herein is in accordance with that proposed by IUB (Nomenclature Committee (1985) Eur J. Biochem., 150, 1-5).

Nucleic Acids

IUB	Meaning
A	A
G	G
C	C
T	T
M	A or C
R	A or G
W	A or T
S	C or G
Y	C or T
K	G or T
V	A or C or G
H	A or C or T
D	A or G or T
B	C or G or T
N	A or G or C or T
.	not A or G or C or T

Amino acids

IUB	3-Letter	Meaning
A	Ala	Alanine
B	Asp, Asn	Aspartic or Asparagine
C	Cys	Cysteine
D	Asp	Aspartic
E	Glu	Glutamic
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
X	Xaa	Unknown
Y	Tyr	Tyrosine
Z	Glu,Gln	Gluatmic or Glutamine
*	End	Stop

The present invention will now be illustrated by the following non-limiting Examples.

Example 1

Hybridisation studies with yeast squalene synthase DNA

The sequences of three enzymes in the sterol biosynthtic pathway have previously been reported for both yeast and mammals (rat or human). In the case of yeast and rat mevalonate kinase, the overall identity is only 32% (Tanaka, R. D., Lee, L. Y., Schafer, B. L., Kratunis, V. J., Mohler, W. A., Robinson, G. W. and Mosley, S. T. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 2872-2876; Oulmouden, A. and Karst, F. (1991) Curr. Genet. 19, 9-14), whilst for farnesyl diphosphate synthase, the homology is 45% identity between yeast and human (Wilkins, D. J., Kutsunai, S. Y. and Edwards, P. A. (1990) J. Biol. Chem. 265, 4607-4614; Anderson, M. S., Yarger, J. G., Burck, C. L. and Poulter, C. D. (1989) J. Biol. Chem. 264, 19176-19184). HMG-CoA reductase displays a higher degree of conservation between yeast and human sequences in the catalytic domain (65%), but the N-terminal membrane-bound portion of the enzyme shows little conservation between yeast and man (Basson, M. E., Thorsness, M., Finer-Moore, J., Stroud, R. M. and Rine, J. (1988) Mol. Cell. Biol. 8, 3797-3808). In none of these cases was one partner cloned by hybridisation to a probe from the other.

We have investigated whether the sequence of human squalene synthase is sufficiently homologous to yeast squalene synthase for human nucleic acid sequences to be detected by cross-hybridisation to yeast DNA probes. Such probes may contain all or part of the yeast DNA sequence encoding squalene synthase, and may contain some or all of a putative isoprenoid binding site that has been identified previously in yeast squalene synthase and other enzymes which bind polyisoprenyl diphosphate molecules (Jennings et al. (1991) as above; Ashby, M. N. and Edwards, P. A. (1990) as above).

Hybridisation of yeast squalene synthase gene probes to yeast and human mRNA

PolyA$^+$ RNA was prepared from yeast strains DB746 and FK5187 and HepG2 cells, and size fractionated on a 1% agarose-formaldehyde gel. The RNA was transferred to Hybond-N membrane and hybridised with a ^{32}P-labelled probe comprising the 440 bp EcoRI-ClaI fragment from the yeast squalene synthase gene, which codes for amino acids 133 - 278, in sodium phosphate/dextran sulphate/Denhardt's buffer at 42°C. The membrane was washed in decreasing concentrations of SSC, 0.1% SDS. Figure 1 shows a typical autoradiograph obtained after washing the membrane at a stringency of 6 x SSC, 0.1% SDS at 50°C. The 1.8 kb yeast squalene synthase mRNA obtained from either the ERG9 or erg9 strain (Lanes 1 and 2 respectively) is clearly revealed, whilst there is no evidence for any signal in the HepG2 track (Lane 3). The EcoRI-ClaI fragment used contains the putative isoprenoid binding site motif in yeast sequalene synthase (amino acids 213 - 239). Similar results were obtained using a 960 bp ClaI-PvuII probe containing sequences encoding the C-terminal 166 amino acids of yeast squalene synthase. Control probes hybridised to single specific bands in the HepG2 track, indicating this RNA is not degraded.

The EcoRI-ClaI yeast probe was also used to screen a human liver cDNA library in the λZap XR vector, in sodium phosphate/dextran sulphate/Denhardt's buffer at 55°C. The membrane was washed in 0.2 x SSC, 0.1% SDS at 55°C. No positive plaques were detected that survived 2 rounds of screening. The same library was subsequently used to isolate human squalene synthase cDNA clones by the PCR-based approach described herein.

Hybridisation of yeast squalene synthase gene probes to human and yeast genomic DNA

Genomic DNA was isolated from human HepG2 cells and yeast strain DB746, digested with restriction enzymes HindIII, EcoRI, EcoRV or BamHI and the products size fractionated on a 0.6% agarose gel. The DNA fragments were transferred to Hybond-N membrane and hybridised with a

1.8kb EcoRI-HpaI fragment from the yeast squalene synthase gene containing the coding sequences for the C-terminal 311 amino acids of yeast squalene synthase in 0.5M sodium phosphate buffer as described above. The membrane was then washed with decreasing concentrations of sodium phosphate, 1% SDS at 60°C. A typical autoradiograph, obtained after washing with 50mM sodium phosphate, 1% SDS is shown in Figure 2. No hybridisation of the yeast probe to human DNA was detected.

Example 2

PCR amplification and cloning of human squalene synthase

Hybridisation experiments of the type described in the Example 1 failed to detect human sequences homologous to yeast squalene synthase. This may in part be due to the differences in preferred codon usage between the two organisms. We have therefore devised a PCR strategy to identify human sequences equivalent to yeast squalene synthase by identifying specific amino acids in the yeast squalene synthase protein sequence which we could predict would be conserved in a human enzyme, even when the surrounding sequence shows little homology.

Both squalene synthase and phytoene synthase catalyse the head-to-head (1'-1) condensation of two allylic isoprenyl diphosphate molecules (farnesyl diphosphate, FPP or geranylgeranyl diphosphate, GGPP) to form the cyclic intermediates presqualene diphosphate, PSPP and prephytoene diphosphate, PPPP respectively (Figure 3). Like squalene synthase, phytoene synthase is a membrane-associated single-chain polypeptide, found in plant chromoplasts and chloroplasts (Dogbo, O., Laferriere, A., D'Harlingue, A. and Camara, B. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 7054-7058; Bramley, P., Teulieres, C., Blain, I., Bird, C. and Schuch, W. (1992) The Plant Journal 2, 343-349; Bartley, G. E., Viitanen, P. V., Bacot, K. O. and Scolnik, P. A. (1992) J. Biol. Chem. 267, 5026-5039), and in certain photosynthetic bacteria (Armstrong, G. A., Alberti, M. and Hearst, J. E. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 9975-9979). In the case of squalene synthase, PSPP undergoes a reductive elimination of a second diphosphate group in a reaction that requires NADPH (Figure 3), whilst the rearrangement of PPPP to phytoene

requires the elimination of a second proton and pyrophosphate anion. We speculated that the apparent mechanistic similarity between the two enzymes might be reflected in a conservation of essential sequences. It has recently been shown that phytoene synthase is encoded by the the crtB genes of two Erwinia species (E. herbicola and E. uredovora) (Sandmann, G. and Misawa, N. (1992) FEMS Microbiol. Letters 90, 253-258), the pys gene in the cyanobacterium Synechoccus PCC7942 (Chamovitz, D., Misawa, N., Sandmann, G. and Hirschberg, J. (1992) FEBS Letters 296, 305-310), and two closely related cDNAs from tomato, Psy1 and pTOM5, which share a low level of homology with the CrtB proteins (Bramley, P., Teulieres, C., Blain, I., Bird, C. and Schuch, W. (1992) The Plant Journal 2, 343-349; Bartley, G. E., Viitanen, P. V., Bacot, K. O. and Scolnik, P. A. (1992) J. Biol. Chem. 267, 5026-5039). Given the apparent mechanistic similarity of the reactions catalysed by the two enzymes, we reasoned that the regions of yeast squalene synthase conserved in phytoene synthases might also be conserved in human squalene synthase.

Comparison of the amino acid sequence of yeast squalene synthase with published sequences of the phytoene synthases revealed three regions of homology (Figure 4). The most C-terminal of these, region III, has previously been described as a possible binding site for the charged head groups of isoprenoid diphosphates (Ashby, M. N. and Edwards, P. A. (1990) as above), although we show herein that there are important differences. We reasoned that the same three conserved regions should also be present in human squalene synthase, even if the overall homology with the corresponding yeast enzyme was low.

We selected two stretches of yeast squalene synthase sequence between amino acids 178-187 in region II and 214-223 in region III, which contain the highest density of residues conserved with phytoene synthase (see Figure 4), and used these to design PCR primers in which the nucleotide sequences were optimised for human codon usage and codon degeneracy minimised. These primers were used in a PCR experiment with cDNA prepared by reverse transcription of polyA$^+$ RNA isolated from the human hepatoma cell line HepG2. HepG2 cells have previously been shown to express readily measurable levels of squalene synthase (Cohen, L. H., Van, M. E. and Griffioen, M. (1989) Biochim. Biophys. Acta 1002, 69-73).

First strand cDNA synthesis was carried out using 1μg of polyA+ mRNA and a cDNA synthesis kit (Phamacia). PCR amplifications were performed in 50mM KCl, 1.5mM HgCl2, 10mM Tris-HCl (pH8.3 at RT) with Taq polymerase for 38 x 1 minute cycles of 94°C (denaturation), 55°C (annealing), 72°C (extension). The degenerate antisense primer A 5′-ATRTTGGTYTTYTGSAGRAASAGGCCCAT-3′ (Figure 4 - region III; see also Figure 6) was used in conjunction with sense primer B, 5′-TAYTGCCAYGTSGCYGGSCTSGTSGG-3′ (Figure 4 - region II; see also Figure 6) on 10ng HepG2 cDNA. Primer A was also used with primer C (5′-TAYCTSATCCTSMGRGCYTAYGA-3′).

The products of the PCR reactions were analysed by electrophoresis on 3% agarose gels. Using primers A and B with HepG2 cDNA or yeast genomic DNA (50 ng), we observed a single product of about 140 bp, which is consistent with the expected spacing between regions II and III in squalene synthase. The 140 bp HepG2-derived PCR fragment was cloned into ddT-tailed Bluescript SKII+ (Holton, T. A. and Graham, M. W. (1991) Nucleic Acids Res. 19, 1156) linearised at the EcoRV site, and sequenced. DNA sequencing showed the cloned fragment contained a single open reading frame extending between the two primers, encoding a novel peptide with limited homology to yeast squalene synthase outside of the primer sequences. This eliminated the possibility that the band we had obtained arose from yeast contamination of the PCR.

PCR-amplified probes were labelled with ^{32}P using random hexamer primers (Feinberg, A. P. and Vogelstein, B. (1983) Anal. Biochem. 132, 6-13), and used to probe an oligodT primed adult human liver cDNA library in the vector λZAP XR (Stratagene). Hybridisations were carried out in 0.28M sodium phosphate (pH7.2), 5 x Denhardt's solution, 10% dextran sulphate, 0.1% SDS at 65°C at probe concentrations of 1 x 10^6 cpm/ml and 1 x 10^7 cpm/membrane. Membranes were washed in 0.5 x SSC, 0.1% SDS at 65°C, and positive plaques visualised by autoradiography. 12 positive lysogens, selected from 1.2 x 10^6 plaques screened, survived three further rounds of screening. Bluescript plasmids were excised from the phage obtained from these lysogens following the manufacturer's protocol, and their cDNA inserts sized by PCR, using primers homologous to the Bluescript flanking sequences. Six of the plasmids were analysed further by restriction mapping and DNA sequencing.

Sequence analysis

DNA sequencing of the selected clones showed that they all belong to a single contiguous cDNA of 2051 bp, with a polyadenylated 3′ end (Figure 5). Within this cDNA there is an open reading frame (ORF) encoding a protein of 417 amino acids (Figure 6). Two of our clones, pHLL6 and pHLL17, contain the whole of the ORF (Figure 5), which begins with an ATG codon at nt 92. The DNA sequence flanking the ATG contains an 8/10 match to a consensus Kozak translation initiation sequence [5′ GCC(A/G)CCATGG 3′] (Herman, R. C. (1989) Trends Biochem. Sci. 14, 219-222). There is an in-frame stop codon 60 nt upstream of the ATG in pHLL17, suggesting that this ATG represents the correct translational start point for the encoded protein. Clones pHLL6 and pHLL17 terminate 400 nt 3′ to the end of the ORF (Figure 5). Clones pHLL2 and pHLL4 contain an additional 300 nt of 3′ untranslated sequence, including two AATAAA polyadenylation signals at nt 1883 and nt 2011 (Manley, J. L. (1988) Biochim. Biophys. Acta 950, 1-12), the latter being separated by 16 nt from a polyA tail.

The ORF contained within pHLL6 and pHLL17 encodes a protein of predicted molecular weight 48,041D, which is in good agreement with the size of full-length rat hepatic squalene synthase, estimated by denaturing gel electrophoresis and Western blotting of a liver microsomal extract (Schechter, I. et al., (1992) as above). The protein sequence shows limited overall homology (42.6% identity) to yeast squalene synthase. However, as shown in Figure 5, the similarities between the two amino acid sequences are clustered into 6 stretches of very high homology (ie >85%), separated by sequences with low identity. Three of these correspond to regions I - III that are also conserved with phytoene synthases, and include the putative isoprenoid binding motif.

Further evidence of the similarity between the protein encoded by our cDNA and yeast squalene synthase is provided by comparison of the hydrophobicity profiles (Figure 7). The human sequence displays four regions of high hydrophobicity. Three of these are located in the half of the protein C-terminal to aa 170, and the overall hydrophobicity plot in this region closely matches that of yeast squalene synthase. In

particular, the human sequence contains a C-terminal hydrophobic tail of sufficient length to span a membrane (21 aa).

Several reports have previously noted the presence of a conserved region in IPP synthases that has been proposed as an allylic isoprenoid binding site (Ashby and Edwards, 1990; Ashby et al., 1992). Figure 8 compares the sequences of this conserved region from FPP-, GGPP- and hexaprenyl-diphosphate synthases with the sequence surrounding region 4 (boxed) of hSQS. Within this region, all enzymes share the conserved motif G(X)3Q(X)6D. There are, however, some significant differences in the homolgies between the squalene synthase/phytoene synthase family and the isopreny diphosphate synthases. Whereas squalene synthase or phytoene synthase conserve an uncharged Asn-Ile pair of amino acids within region 4, the isoprenyl dishosphate synthases contain a pair of acidic Asp residues (positions 10 and 11 in Figure 8), the first of which is reported to be essential for activity in rat FPP synthase (Ashby et al., 1992) . Further differences are also apparent, both within region 4 at positions 13 and 18, and in the adjacent C-terminal sequence. Therefore, we suggest that the role of this region in the polyisoprene synthases and isoprenyl diphosphate synthases is functionally distinct, and reflects the different nature of the condensation reactions (1'-1 vs 1'-4) catalysed by the two types of enzyme.

Example 3

Expression of human squalene synthase in E coli

In order to confirm the identity of our cDNA, we undertook expression experiments in E. coli, which lacks squalene synthase. The cDNA insert from pHLL6 was subcloned into the E. coli expression vector pTB361, which carries a phage T7 promoter and T4 terminator, separated by a polylinker region containing a Shine-Dalgarno sequence and an ATG codon (Figure 9). In one construct, a 1.65kb EcoNI-XhoI fragment from pHLL6 was inserted between the NdeI and XhoI sites in pTB361, affording pHSQS6. pHSQS6 contains the complete ORF from the human cDNA, but with an additional 8 nt inserted between the

Shine-Dalgarno sequence and the ATG. In a second construct, the StyI-XhoI fragment from pHLL6, which lacks the first six amino acids of the ORF, was cloned between the BglII and XhoI sites of pTB361, producing pHSQS7. This construct retains the optimal spacing between the Shine-Dalgarno and ATG codons in pTB361, but exchanges the first six amino acids of the human ORF (MEFVKC) for four vector-encoded amino acids (MVPD) (Figure 9).

Plasmid pHSQS6 was constructed by linearising pHLL6 with EcoNI, filling in the 5′ ends with Klenow polymerase. Following digestion with XhoI, the 1.65kb EcoNI′ - XhoI fragment was subcloned into pTB361 which had been digested with NdeI, blunt-ended using Klenow polymerase, and then cut with XhoI. To construct pHSQS7, pHLL6 was first digested with StyI, the overhanging 5′ end partially filled in with Klenow polymerase + dCTP and dATP, and the product further digested with XhoI. The 1.65kb StyI′ - XhoI fragment was ligated between the partially filled-in BglII site (Klenow polymerase + dGTP, dATP and dTTP) and XhoI site in pTB361.

Both pHSQSS6 and pHSQS7 were transformed into E. coli strain BL21(DE3), which contains the T7 RNA polymerase gene under control of the lac uv5 promoter (Studier, F. W. and Moffatt, B. A. (1986) as above). Transformants carrying either pHSQS6, pHSQS7 or the parental vector were cultured in the presence or absence of IPTG, and cell extracts were prepared. 100ml cultures of E. coli BL21(DE3) containing the appropriate plasmid were grown to O.D600 1.0 in Lbroth containing tetracycline at 10µg/ml. The cells were induced with 0.1mM IPTG for 3 hours and then harvested at 2000g (max). The pellets were washed twice with ice-cold 50mM potassium phosphate, 1mM EDTA, (pH7.4) and resuspended in 1ml ice-cold homogenisation buffer (50mM potassium phosphate, 1mM EDTA, 0.5mM DTT, pH7.4). Cells were disrupted by sonication, cooling on ice between pulses. The homogenate was centrifuged at 2,000g for 5 minutes at 4°C. The 2,000g supernatant was further centrifuged at 12,000g for 15 minutes at 4°C, and the supernatants stored in aliquots at -70°C. Rat liver microsomes (100,000g particulate fraction) were prepared according to published procedures (Slakey, L. L., Craig, M. C., Beytia, E., Briedis, A.,

Feldbruegge, D. H., Dugan, R. E., Qureshi, A. A., Subbarayan, C. and Porter, J. W. (1972) J. Biol. Chem. 247, 3014-3022).

The extracts from cells containing pHSQS6 or pHSQS7 were assayed for squalene synthase activity. Squalene synthase catalyses the elimination of the 1-pro-S hydrogen from one molecule of FPP molecule to the solvent, which is subsequently replaced by a hydrogen donated from NADPH (Agnew, W. S. and Popjak, G. (1978) J. Biol. Chem. 253, 4566-4573). The reaction can therefore be followed by monitoring release of ^{3}H from 1-[^{3}H]FPP into the solvent or incorporation into squalene. 12,000g *E. coli* extracts or rat microsomes containing 20µg of total protein were incubated in 200ml buffer containing 50mM potassium phosphate (pH7.4), 5mM MgCl2, 10mM KF, 0.9mM NADPH, 20mM 1-[^{3}H]-FPP (0.1mCi) for 15 minutes at 37°C. The reaction was stopped by addition of KOH to 0.8% and the products fractionated on an Analytichem Bond Elute C18 column (Varian). The column was washed with 0.1M KOH to recover the aqueous fraction (Fraction 1), and the squalene-containing fraction (Fraction 2) was subsequently eluted with 10% ethyl acetate in hexane. Unreacted FPP was retained on the column. The eluted fractions were analysed by scintillation counting.

The identity of squalene in Fraction 2 was confirmed by HPLC analysis. Material from Fraction 2 was loaded onto a Spherisorb S30DS1 column (Hewlett Packard) and samples eluted in a linear gradient from 60% acetonitrile:40% water to 95% acetonitrile:5% propan-2-ol for 15 minutes, followed by a second linear gradient from 95:5 acetonitrile:propan-2-ol to 80% acetonitrile:20% propan-2-ol for 25 minutes. In this system, squalene has a retention time of 20-21 minutes. Column fractions were collected for 30 seconds, monitoring the UV absorbance at 210 nm, and analysed by scintillation counting.

The assay results are summarised in Table 1. BL21(DE3) carrying the parental vector pTB361 contained no detectable squalene synthase activity. However, extracts prepared from IPTG-induced cells containing pHSQS6 showed a low level of squalene synthase activity, as measured by the appearance of radioactivity in the both fractions. The levels of enzyme were dramatically increased in extracts prepared from both uninduced and induced cells containing pHSQS7. Figure 10 shows the HPLC trace of Fraction 2 from cells containing pHSQS7. The

only significant component of this fraction has a retention time on the column identical with squalene, and coelutes with a peak in the UV absorbance spectrum. The significant level of squalene synthase activity detected in extracts from uninduced cells containing this plasmid, is presumably due to a background level of T7 polymerase caused to leaky transcription from the lac uv5 promoter in the host strain (Studier, F. W. and Hoffatt, B. A. (1986) as above). Cells containing pHSQS7 grew slowly even in the absence of IPTG, and the low induction ratio probably reflects toxic effects due to overproduction of squalene synthase. Deletion of DNA sequences specifying the C-terminal 85 amino acids from pHSQS6 or pHSQS7 abolished the production of squalene synthase activity in the extracts. These results confirm that the cDNA described herewithin encodes human squalene synthase.

TABLE 1. EXPRESSION OF HUMAN SQUALENE SYNTHASE IN E. coli

Plasmid	IPTG	Specific activity
pTB361	-	0
	+	0
pHSQS6	-	0
	+	0.02 (0.01)
pHSQS7	-	0.57 (0.07)
	+	1.01 (0.01)

Specific activity of squalene synthase in nmoles squalene/min/mg protein calculated. Figures in parentheses indicate the range from duplicate determinations. In this assay, rat liver microsomes gave a specific activity of 0.73(+/-0.01) n moles squalene/min/mg protein, calculated from fraction 2.

Using 1-[^{3}H] FPP labelled equally at the 1-pro-R and 1-pro-S positions as a substrate, the ratio of ^{3}H incorporated into squalene to 3H released into the solvent is expected to be 3:1. In practice, the measured ratio is slightly less than the theoretical figure due to water-soluble degradation products contaminating the radiolabelled FPP. In microsomal preparations the squalene produced can also be further metabolised by other membrane bound enzymes in the sterol pathway. We observed a [^{3}H(Fraction 2):^{3}H(Fraction 1)] ratio of 2.4:1 with E. coli extracts and 2.1:1 with rat hepatocyte microsomes.

The specific activity of squalene synthase obtained with pHSQS7 in E. coli is comparable to that found in rat hepatic microsomes. Although the cDNA sequences contained within pHSQS6 and pHSQS7 encode full-length squalene synthase, cell fractionation experiments showed that 90% of the activity was present in the soluble fraction, and only 10% was recovered in the P_{100} pellet, suggesting that membrane localisation is

inefficient in E. coli. Similar results were obtained by Fegueur et al. (Fegueur, M. et al (1991) as above) when they expressed yeast squalene synthase in E. coli. This soluble activity could be due to proteolysis of the recombinant squalene synthase. Alternatively, the failure of the enzyme to efficiently associate with the cell membrane may be due to the lack of a suitable transport pathway in E. coli, or the absence of downstream membrane-localised enzymes such as squalene oxidase which may interact with squalene synthase in a multi-enzyme complex. However, the formation of solubilsed enzyme in this host may facilitate the overexpression of recombinant squalene synthase for structural studies, and provide a source of enzyme in a non-sterol metabolising host for enzymatic studies or the screening of compounds to identify inhibitors of squalene synthase.

SEQ ID NO:1:

SEQUENCE CHARACTERISTICS
(A) LENGTH:2052 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

SEQUENCE DESCRIPTION:SEQ ID NO:1:

```
CGAGACCTAC TCCACAGGTC CAGCCGGCCG GTGAGCGCCT GGGGACCGCA GAGGTGAGAG    60

TCGCGCCCGG GAGTCCGCCG CCTGCGCCAG G ATG GAG TTC GTG AAA TGC CTT       112
                                    Met Glu Phe Val Lys Cys Leu
                                    1               5

GGC CAC CCC GAA GAG TTC TAC AAC CTG GTG CGC TTC CGG ATC GGG GGC      160
Gly His Pro Glu Glu Phe Tyr Asn Leu Val Arg Phe Arg Ile Gly Gly
        10                  15                  20
```

```
AAG CGG AAG GTG ATG CCC AAG ATG GAC CAG GAC TCG CTC AGC AGC AGC      208
Lys Arg Lys Val Met Pro Lys Met Asp Gln Asp Ser Leu Ser Ser Ser
    25                  30                  35

CTG AAA ACT TGC TAC AAG TAT CTC AAT CAG ACC AGT CGC AGT TTC GCA      256
Leu Lys Thr Cys Tyr Lys Tyr Leu Asn Gln Thr Ser Arg Ser Phe Ala
40                  45                  50                  55

GCT GTT ATC CAG GCG CTG GAT GGG GAA ATG CGC AAC GCA GTG TGC ATA      304
Ala Val Ile Gln Ala Leu Asp Gly Glu Met Arg Asn Ala Val Cys Ile
                60                  65                  70

TTT TAT CTG GTT CTC CGA GCT CTG GAC ACA CTG GAA GAT GAC ATG ACC      352
Phe Tyr Leu Val Leu Arg Ala Leu Asp Thr Leu Glu Asp Asp Met Thr
            75                  80                  85

ATC AGT GTG GAA AAG AAG GTC CCG CTG TTA CAC AAC TTT CAC TCT TTC      400
Ile Ser Val Glu Lys Lys Val Pro Leu Leu His Asn Phe His Ser Phe
        90                  95                  100

CTT TAC CAA CCA GAC TGG CGG TTC ATG GAG AGC AAG GAG AAG GAT CGC      448
Leu Tyr Gln Pro Asp Trp Arg Phe Met Glu Ser Lys Glu Lys Asp Arg
    105                 110                 115

CAG GTG CTG GAG GAC TTC CCA ACG ATC TCC CTT GAG TTT AGA AAT CTG      496
Gln Val Leu Glu Asp Phe Pro Thr Ile Ser Leu Glu Phe Arg Asn Leu
120                 125                 130                 135

GCT GAG AAA TAC CAA ACA GTG ATT GCC GAC ATT TGC CGG AGA ATG GGC      544
Ala Glu Lys Tyr Gln Thr Val Ile Ala Asp Ile Cys Arg Arg Met Gly
                140                 145                 150

ATT GGG ATG GCA GAG TTT TTG GAT AAG CAT GTG ACC TCT GAA CAG GAG      592
Ile Gly Met Ala Glu Phe Leu Asp Lys His Val Thr Ser Glu Gln Glu
            155                 160                 165
```

```
TGG GAC AAG TAC TGC CAC TAT GTT GCT GGG CTG GTC GGA ATT GGC CTT      640
Trp Asp Lys Tyr Cys His Tyr Val Ala Gly Leu Val Gly Ile Gly Leu
        170                 175                 180

TCC CGT CTT TTC TCA GCC TCA GAG TTT GAA GAC CCC TTA GTT GGT GAA      688
Ser Arg Leu Phe Ser Ala Ser Glu Phe Glu Asp Pro Leu Val Gly Glu
    185                 190                 195

GAT ACA GAA CGT GCC AAC TCT ATG GGC CTG TTT CTG CAG AAA ACA AAC      736
Asp Thr Glu Arg Ala Asn Ser Met Gly Leu Phe Leu Gln Lys Thr Asn
200                 205                 210                 215

ATC ATC CGT GAC TAT CTG GAA GAC CAG cAA GGA GGA AGA GAG TTC TGG      784
Ile Ile Arg Asp Tyr Leu Glu Asp Gln Gln Gly Gly Arg Glu Phe Trp
                220                 225                 230

CCT CAA GAG GTT TGG AGC AGG TAT GTT AAG AAG TTA GGG GAT TTT GCT      832
Pro Gln Glu Val Trp Ser Arg Tyr Val Lys Lys Leu Gly Asp Phe Ala
            235                 240                 245

AAG CCG GAG AAT ATT GAC TTG GCC GTG CAG TGC CTG AAT GAA CTT ATA      880
Lys Pro Glu Asn Ile Asp Leu Ala Val Gln Cys Leu Asn Glu Leu Ile
        250                 255                 260

ACC AAT GCA CTG CAC CAC atC CCA GAT GTC ATC ACC TAC CTT TCG AGA      928
Thr Asn Ala Leu His His Ile Pro Asp Val Ile Thr Tyr Leu Ser Arg
    265                 270                 275

CTC AGA AAC CAG AGT GTG TTT AAC TTC TGT GCT ATT CCA CAG GTG ATG      976
Leu Arg Asn Gln Ser Val Phe Asn Phe Cys Ala Ile Pro Gln Val Met
280                 285                 290                 295

GCC ATT GCC ACT TTG GCT GCC TGT TAT AAT AAC CAG CAG GTG TTC AAA     1024
Ala Ile Ala Thr Leu Ala Ala Cys Tyr Asn Asn Gln Gln Val Phe Lys
                300                 305                 310
```

GGG GCA GTG AAG ATT CGG AAA GGG CAA GCA GTG ACC CTG ATG ATG GAT 1072
Gly Ala Val Lys Ile Arg Lys Gly Gln Ala Val Thr Leu Met Met Asp
315 320 325

GCC ACC AAT ATG CCA GCT GTC AAA GCC ATC ATA TAT CAG TAT ATG GAA 1120
Ala Thr Asn Met Pro Ala Val Lys Ala Ile Ile Tyr Gln Tyr Met Glu
330 335 340

GAG ATT TAT CAT AGA ATC CCC GAC TCA GAC CCA TCT TCT AGC AAA ACA 1168
Glu Ile Tyr His Arg Ile Pro Asp Ser Asp Pro Ser Ser Ser Lys Thr
345 350 355

AGG CAG ATC ATC TCC ACC ATC CGG ACG CAG AAT CTT CCC AAC TGT CAG 1216
Arg Gln Ile Ile Ser Thr Ile Arg Thr Gln Asn Leu Pro Asn Cys Gln
360 365 370 375

CTG ATT TCC CGA AGC CAC TAC TCC CCC ATC TAC CTG TCG TTT GTC ATG 1264
Leu Ile Ser Arg Ser His Tyr Ser Pro Ile Tyr Leu Ser Phe Val Met
380 385 390

CTT TTG GCT GCC CTG AGC TGG CAG TAC CTG GCC ACT CTC TCC CAG GTA 1312
Leu Leu Ala Ala Leu Ser Trp Gln Tyr Leu Ala Thr Leu Ser Gln Val
395 400 305

ACA GAA GAC TAT GTT CAG ACT GGA GAA CAC TGA TCCCAAATTT GTCCATAGCT 1365
Thr Glu Asp Tyr Val Gln Thr Gly Glu His End
410 415

GAAGTCCACC ATAAAGTGGA TTTACTTTTT TTCTTTAAGG ATGGATGTTG TGTTCTCTTT 1425

ATTTTTTTCC TACTACTTTA ATCCCTAAAA GAACGCTGTG TGGCTGGGAC CTTTAGGAAA 1485

GTGAAATGCA GGTGAGAAGA ACCTAAACAT GAAAGGAAAG GGTGCCTCAT CCCAGCAACC 1545

TGTCCTTGTG GGTGATGATC ACTGTGCTGC TTGTGGCTCA TGGCAGAGCA TTCAGTGCCA 1605

CGGTTTAGGT GAAGTCGCTG CATATGTGAC TGTCATGAGA TCCTACTTAG TATGATCCTG 1665

GCTAGAATGA TAATTAAAAG TATTTAATTT GAAGCACCAT TTGAATGTTC GTAATAGTAG 1725

AAAATGATGT GAATTTTCTT TCTGTTCGGC TCCTATTTTT CTCATCATTT TGTTTTCTTT 1785

AATTGGGTTG AATGGAGTAG ATAGAAATAT TTATGGTTTA GGTAACAGTT AGATGTTTCC 1845

TAAGAATGCA AACTGCCTTT TCCACACAAA GGCTGGGAAT AAAATTCTGG GTATTCTCGT 1905

ATTCTCATTT AAAGGAGTTT AGCTTTCAGA GAGAAACAGC AGGATTGCTT TTGACCTTTT 1965

AGAAGATTGG TCTCCAGTAA AGGTGGACAT TTTTGAGATT TTTATAATAA AGAATTTAAT 2025

TGCTCTGCAA AAAAAAAAAA AAAAAA 2051

SEQ ID NO:2:

SEQUENCE CHARACTERISTICS:
(A) LENGTH:417 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

SEQUENCE DESCRIPTION:SEQ ID NO:2:

```
Met Glu Phe Val Lys Cys Leu Gly His Pro Glu Glu Phe Tyr Asn
1               5                  10                  15

Leu Val Arg Phe Arg Ile Gly Gly Lys Arg Lys Val Met Pro Lys
                20                  25                  30

Met Asp Gln Asp Ser Leu Ser Ser Ser Leu Lys Thr Cys Tyr Lys
                35                  40                  45

Tyr Leu Asn Gln Thr Ser Arg Ser Phe Ala Ala Val Ile Gln Ala
                50                  55                  60

Leu Asp Gly Glu Met Arg Asn Ala Val Cys Ile Phe Tyr Leu Val
                65                  70                  75

Leu Arg Ala Leu Asp Thr Leu Glu Asp Asp Met Thr Ile Ser Val
                80                  85                  90

Glu Lys Lys Val Pro Leu Leu His Asn Phe His Ser Phe Leu Tyr
                95                  100                 105

Gln Pro Asp Trp Arg Phe Met Glu Ser Lys Glu Lys Asp Arg Gln
                110                 115                 120
```

Val Leu Glu Asp Phe Pro Thr Ile Ser Leu Glu Phe Arg Asn Leu
125 130 135

Ala Glu Lys Tyr Gln Thr Val Ile Ala Asp Ile Cys Arg Arg Met
140 145 150

Gly Ile Gly Met Ala Glu Phe Leu Asp Lys His Val Thr Ser Glu
155 160 165

Gln Glu Trp Asp Lys Tyr Cys His Tyr Val Ala Gly Leu Val Gly
170 175 180

Ile Gly Leu Ser Arg Leu Phe Ser Ala Ser Glu Phe Glu Asp Pro
185 190 195

Leu Val Gly Glu Asp Thr Glu Arg Ala Asn Ser Met Gly Leu Phe
200 205 210

Leu Gln Lys Thr Asn Ile Ile Arg Asp Tyr Leu Glu Asp Gln Gln
215 220 225

Gly Gly Arg Glu Phe Trp Pro Gln Glu Val Trp Ser Arg Tyr Val
230 235 240

Lys Lys Leu Gly Asp Phe Ala Lys Pro Glu Asn Ile Asp Leu Ala
245 250 255

Val Gln Cys Leu Asn Glu Leu Ile Thr Asn Ala Leu His His Ile
260 265 270

Pro Asp Val Ile Thr Tyr Leu Ser Arg Leu Arg Asn Gln Ser Val
275 280 285

Phe Asn Phe Cys Ala Ile Pro Gln Val Met Ala Ile Ala Thr Leu
290 295 300

Ala Ala Cys Tyr Asn Asn Gln Gln Val Phe Lys Gly Ala Val Lys
305 310 315

Ile Arg Lys Gly Gln Ala Val Thr Leu Met Met Asp Ala Thr Asn
320 325 330

Met Pro Ala Val Lys Ala Ile Ile Tyr Gln Tyr Met Glu Glu Ile
335 340 345

Tyr His Arg Ile Pro Asp Ser Asp Pro Ser Ser Ser Lys Thr Arg
350 355 360

Gln Ile Ile Ser Thr Ile Arg Thr Gln Asn Leu Pro Asn Cys Gln
365 370 375

Leu Ile Ser Arg Ser His Tyr Ser Pro Ile Tyr Leu Ser Phe Val
380 385 390

Met Leu Leu Ala Ala Leu Ser Trp Gln Tyr Leu Ala Thr Leu Ser
395 400 405

Gln Val Thr Glu Asp Tyr Val Gln Thr Gly Glu His End
410 415

CLAIMS

1. A polypeptide which has all or part of the amino acid sequence and one or more of the biological properties typical of naturally occurring human squalene synthase.

2. A polypeptide having biological properties characteristic of and all or part of the amino acid sequence of human squalene synthase.

3. A polypeptide having substantially the amino acid sequence designated SEQ ID No 2, which polypeptide is free from association with human protein.

4. A polypeptide as claimed in claim 1, 2 or 3 wherein the polypeptide has the amino acid sequence designated SEQ ID No 2.

5. A DNA sequence coding for all or part of a squalene synthase as defined in any one of the preceeding claims.

6. A DNA sequence as claimed in claim 5 which is selected from:

(a) the DNA sequence set forth in SEQ ID No 1 or its complementary strand;

(b) a DNA sequence which hybridises to the DNA sequences in SEQ ID No 1 or to fragments thereof; and

(c) a DNA sequence which, but for the degeneracy of the genetic code, would hybridize to the DNA sequence in SEQ ID No 1.

7. A recombinant vector containing a DNA sequence as defined in claim 5, 6 or 7.

8. A process for the preparation of a recombinant vector as defined in claim 7 which comprises inserting a DNA sequence as defined in claim 5, 6 or 7 into a vector.

9. A procaryotic or eucaryotic host cell stably transformed or transfected with a recombinant vector as defined in claim 7.

10. A process for the preparation of a procaryotic or eucaryotic host cell as defined in claim 9 which comprises transforming or transfecting a procaryotic or eucaryotic cell with a recombinant vector as defined in claim 7 whereby to yield a stably transformed or transfected procaryotic or eucaryotic host.

11. A process for the preparation of a squalene synthase as defined in claim 1, 2 or 3 which comprises culturing a procaryotic or eucaryotic host cell as claimed in claim 9 whereby to obtain said derivative.

12. A process as claimed in claim 11 wherein the process includes the step of isolating the said derivative produced by expression of the DNA sequence of the invention in the recombinant vector.

13. The use of a recombinant squalene synthase as defined in claim 1, 2 or 3 to detect inhibitors of human squalene synthase.

14. The use as claimed in claim 13 wherein the method comprises the preparation of cell extracts containing particulate or soluble cell material from a procaryotic or eucaryotic host producing a squalene synthase by expression of all or part of a human squalene synthase gene as defined in claim 4, 5 or 6, and the use of such extracts to measure the inhibitory effects of compounds on the activity of the squalene synthase.

15. A polypeptide when prepared by the process claimed in claim 10.

16. A polypeptide substantially as hereinbefore described.

17. A vector substantially as hereinbefore described.

18. A procaryotic or eucaryotic host cell substantially as hereinbefore described.

19. A DNA sequence coding for all or part of a squalene synthase substantially as hereinbefore described.

20. A process for the preparation of a procaryotic or eucaryotic host cell substantially as hereinbefore described.

21. A process for the preparation of a squalene synthase substantially as hereinbefore described.

22. The use of a recombinant squalene synthase to detect inhibitors of human squalene synthase substantially as hereinbefore described.

Patents Act 1977 Examiner's report to the Comptroller under Section 17 (The Search report) -42-	Application number GB 9323035.7
Relevant Technical Fields (i) UK Cl (Ed.M) C3H (HB7E) (ii) Int Cl (Ed.5) C12N (9/10, 9/12, 15/54)	Search Examiner DR D ELSY Date of completion of Search 10 FEBRUARY 1994
Databases (see below) (i) UK Patent Office collections of GB, EP, WO and US patent specifications. (ii) ONLINE DATABASES: WPI, DIALOG (BIOTECH), CAS ONLINE	Documents considered relevant following a search in respect of Claims :- 1-22

Categories of documents

X:	Document indicating lack of novelty or of inventive step.	P:	Document published on or after the declared priority date but before the filing date of the present application.
Y:	Document indicating lack of inventive step if combined with one or more other documents of the same category.	E:	Patent document published on or after, but with priority date earlier than, the filing date of the present application.
A:	Document indicating technological background and/or state of the art.	&:	Member of the same patent family; corresponding document.

Category	Identity of document and relevant passages	Relevant to claim(s)
X	GB 2249099 A (SQUIBB) see claims	1-22
P X	Chemical Abstracts 119(17): 176553X & J Biol-Chem, Vol 268(17), pages 12818-12824, (1993) (JIANG G ET AL)	1-22
P X	Biosis Abstract Number: 9601574, & Mol Cell Biol, 13(5), (1993), pages 2706-2717 (ROBINSON G W ET AL)	1-22
X	Biosis Number: 95004297, & J Biol Chem, Vol 267 (30), (1992) pages 21368-21374	1-22

Databases:The UK Patent Office database comprises classified collections of GB, EP, WO and US patent specifications as outlined periodically in the Official Journal (Patents). The on-line databases considered for search are also listed periodically in the Official Journal (Patents).